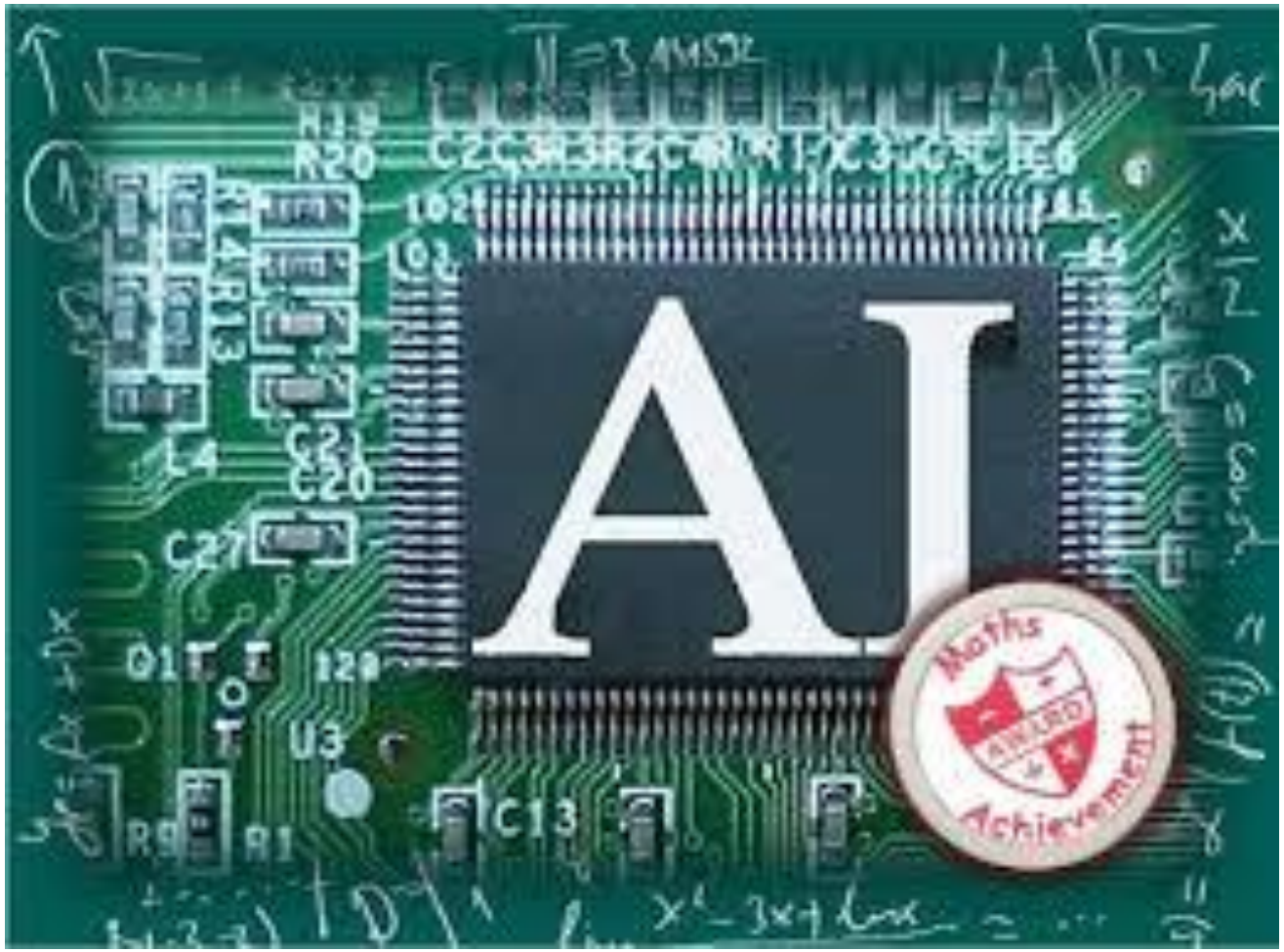
METODI MATEMATICI PER L'INTELLIGENZA ARTIFICIALE

LEZIONE 0

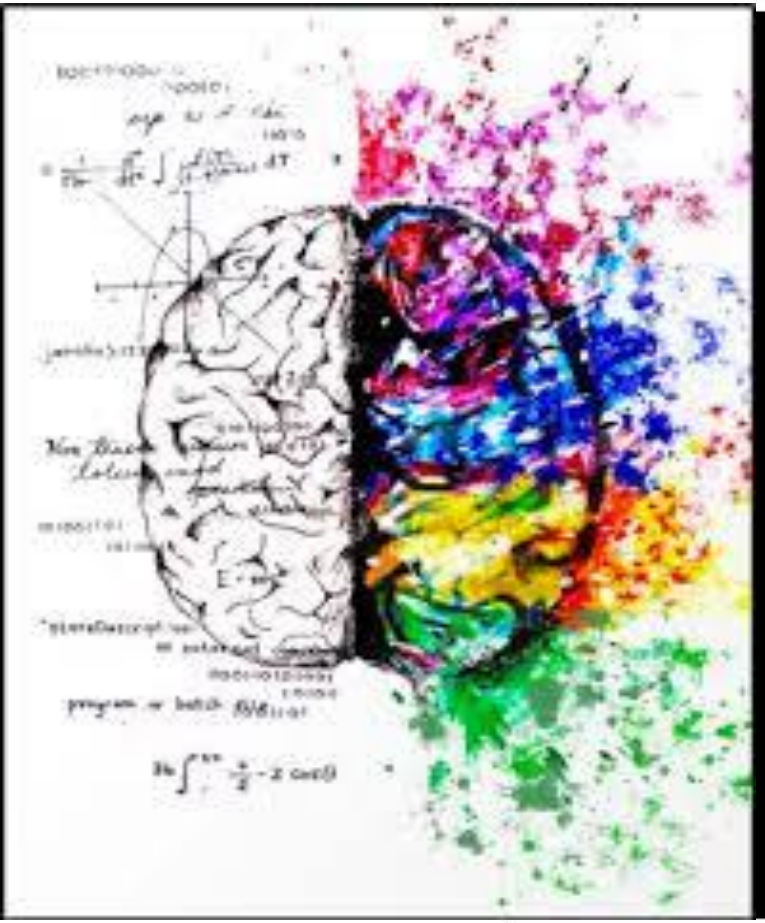


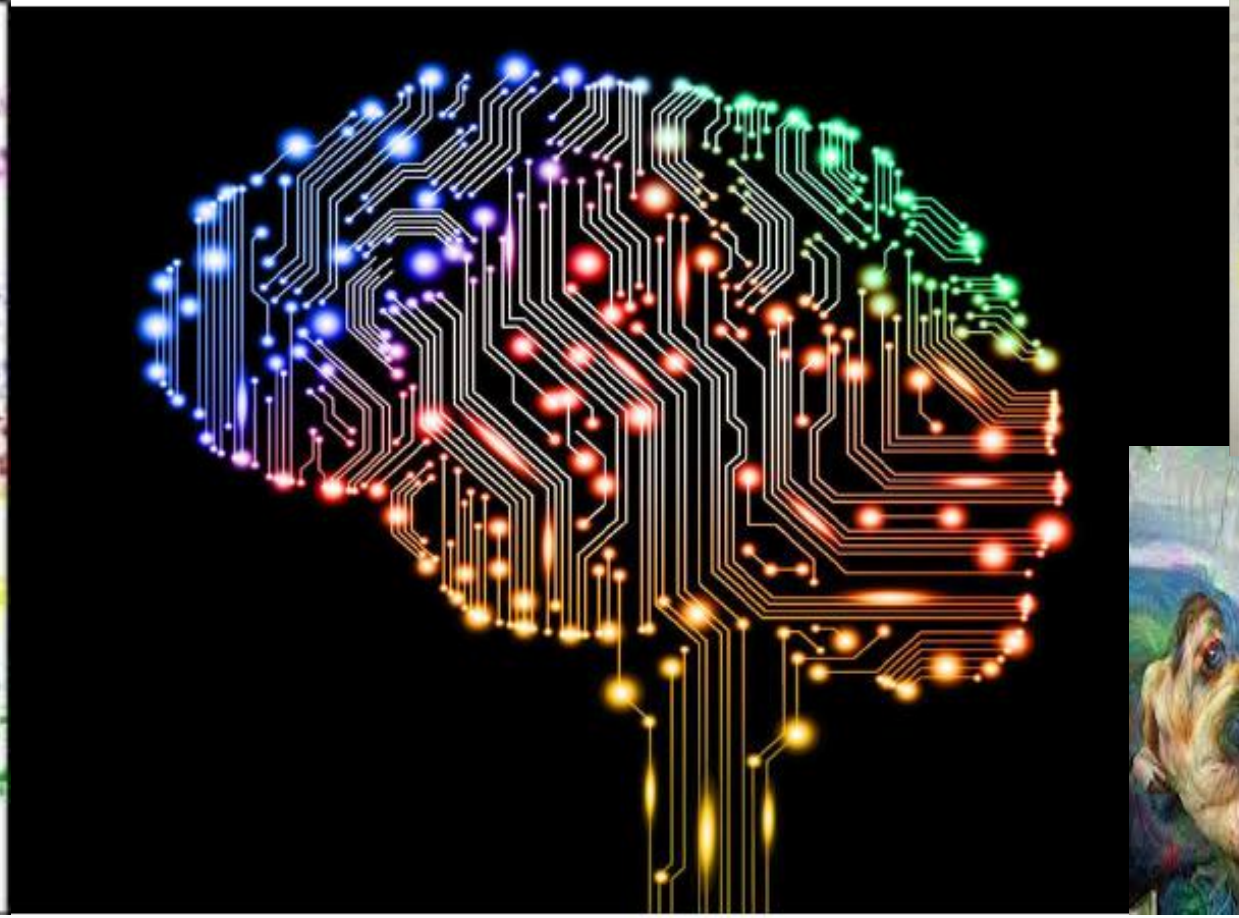
Adriano Barra
Dipartimento di Matematica e Fisica «Ennio De Giorgi», Stanza 455
adriano.barra@gmail.com adriano.barra@unisalento.it
www.adrianobarra.com (teaching)

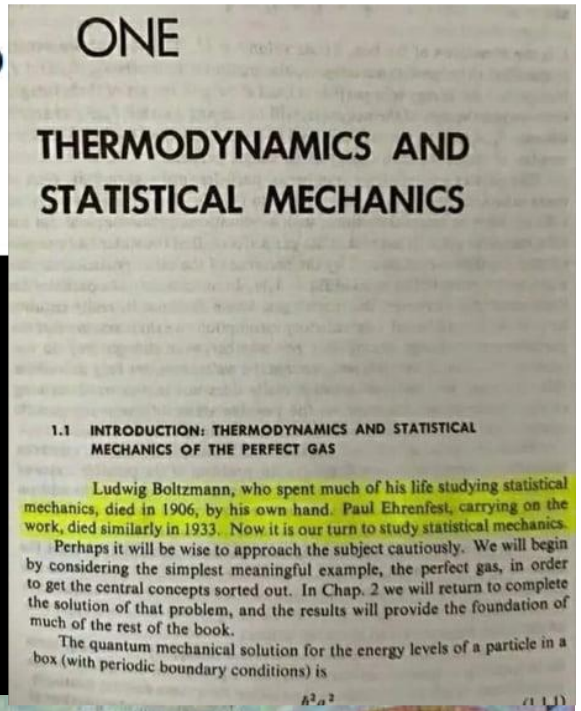
CORSO DI NEURAL NETWORKS & MACHINE LEARNING

Adriano Barra
adriano.barra@gmail.com, adriano.barra@unisalento.it
www.adrianobarra.com Stanza 455 (Fiorini)

Lezione Uno







ONE

THERMODYNAMICS AND STATISTICAL MECHANICS

1.1 INTRODUCTION: THERMODYNAMICS AND STATISTICAL MECHANICS OF THE PERFECT GAS

Ludwig Boltzmann, who spent much of his life studying statistical mechanics, died in 1906, by his own hand. Paul Ehrenfest, carrying on the work, died similarly in 1933. Now it is our turn to study statistical mechanics.

Perhaps it will be wise to approach the subject cautiously. We will begin by considering the simplest meaningful example, the perfect gas, in order to get the central concepts sorted out. In Chap. 2 we will return to complete the solution of that problem, and the results will provide the foundation of much of the rest of the book.

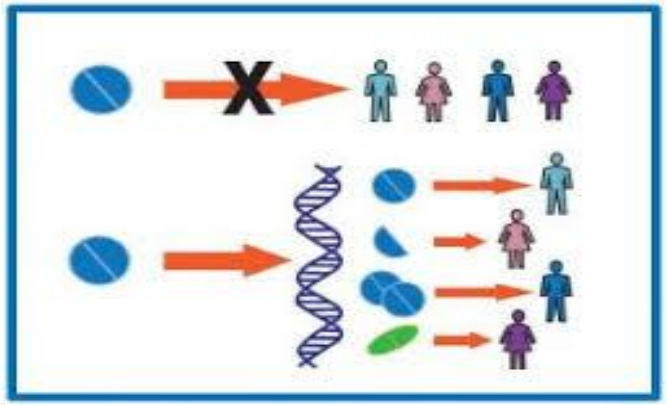
The quantum mechanical solution for the energy levels of a particle in a box (with periodic boundary conditions) is

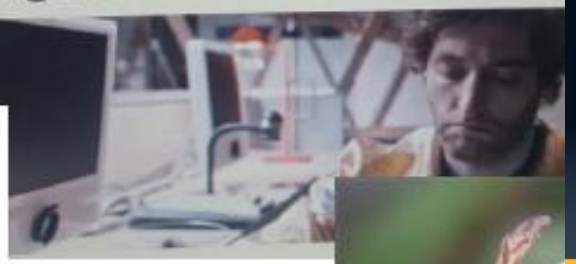
Uno sguardo alle reti neuronali da una prospettiva di meccanica statistica

Adriano Barra[1]

[1] *Dipartimento di Matematica e Fisica Ennio De Giorgi, Universita' del Salento*

(Dated: January 31, 2017)





THE TIMES OF INDIA

Blogs

Home Blogs Times View Times Impact City India World Entertainment Tech Sports Lifestyle Environment Science

News » Blogs » Tech Blogs » Are we ready to face the 'deep learning' revolution?

Are we ready to face the 'deep learning' revolution?

February 8, 2017, 7:49 PM IST Dr Deepak Garg in Breaking Shackles | Tech | TOI

TECH

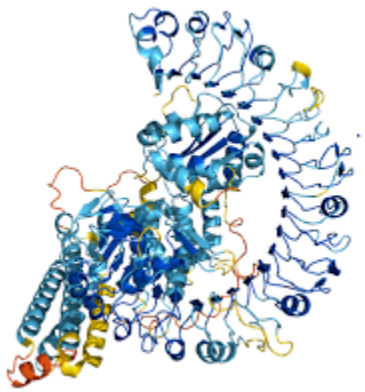
AI Solves 50-Year-Old Biology 'Grand Challenge' Decades Before Experts Predicted

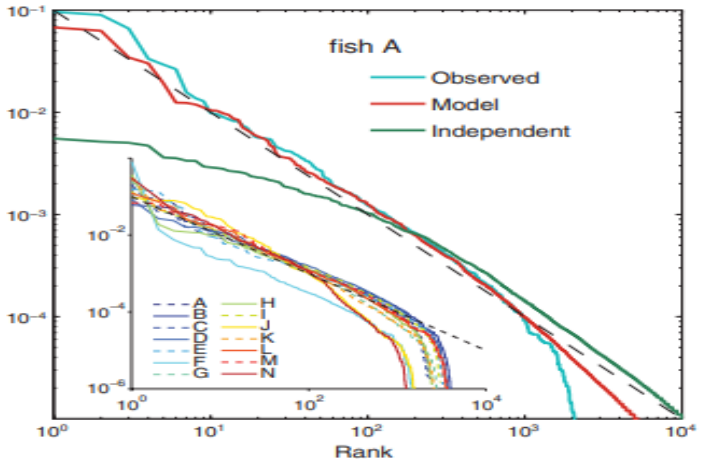
PETER DOCKRILL 1 DECEMBER 2020

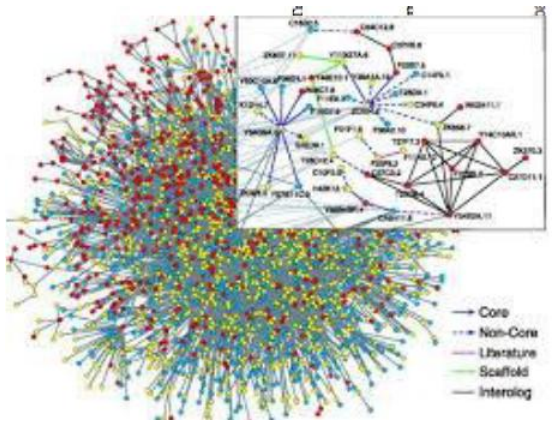
Cosa sta succedendo?
-è esplosa la tecnologia: abbiamo finalmente una rappresentazione digitale della «realtà»

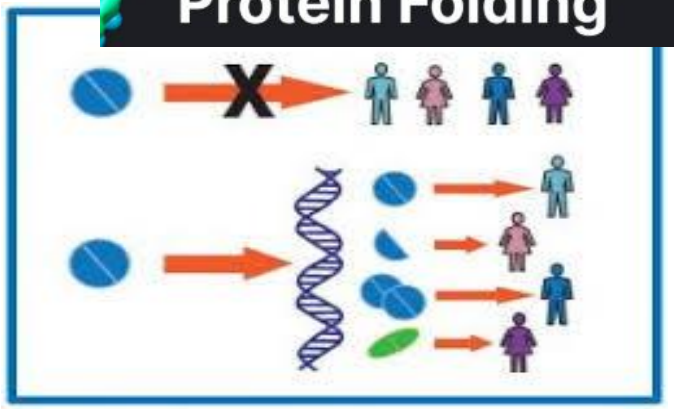
-stiamo collezionando molto più «del necessario x il training», ora abbiamo high D data
e.g. per health-care (inabbracciabile senza l'inferenza statistica ad altà dimensionalità)

-stiamo costruendo l'inferenza statistica ad altà dimensionalità, di cui una rappresentazione è il Machine Learning che si implementa nelle Neural Networks





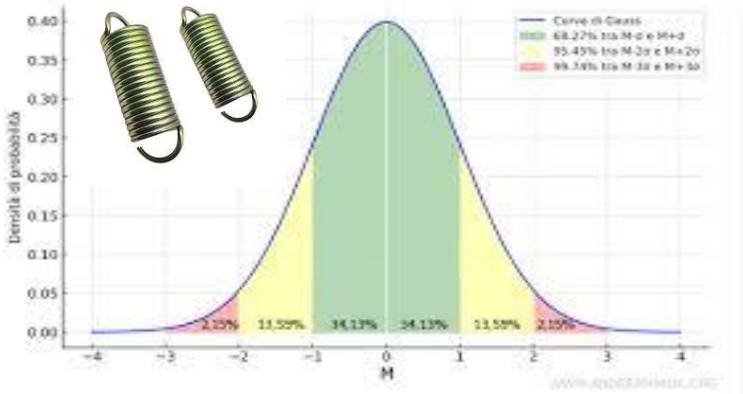




Il riduzionismo va a braccetto con l'inferenza statistica a bassa dimensionalità quindi per costruire l'Ingegneria del previo secolo era perfetto.

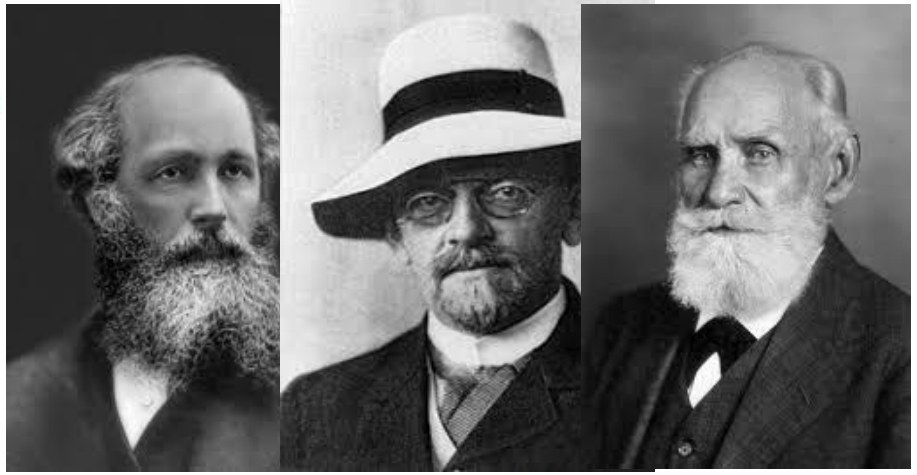
Superarlo è stato –ed è- di proibitiva difficoltà (e.g. chaos, RSB, non unicità,etc e questo approccio sembra per la prima volta una via percorribile.

OGNI SECOLO SI DOTA DELLA STATISTICA DI CUI HA BISOGNO

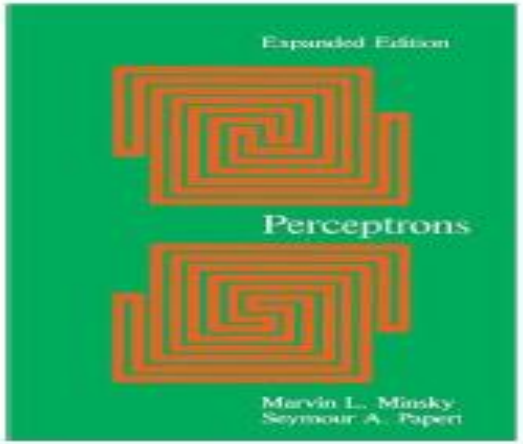


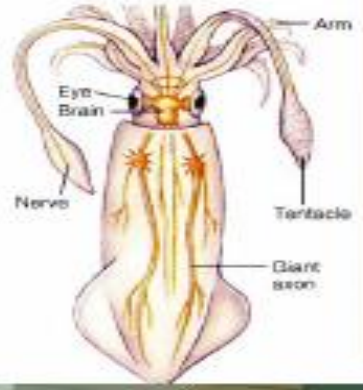
Le reti neurali negli ultimi 100 anni (il passato remoto lo salto qui ma esiste, e.g. Babbage machine...)

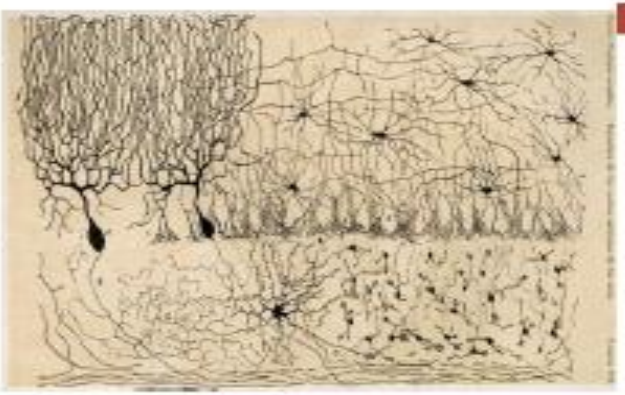
- Ivan Pavlov 1849-1936
- Santiago R.Y. Cajal 1879-1936
- Donald Hebb 1904-1985
- Alan L. Hodking 1914-1998
- Andrew Huxley 1917-2012
- Marvin L. Minsky 1927-2016
- Seymour Papert 1928-2016
- Frank Rosemblatt 1928-1971

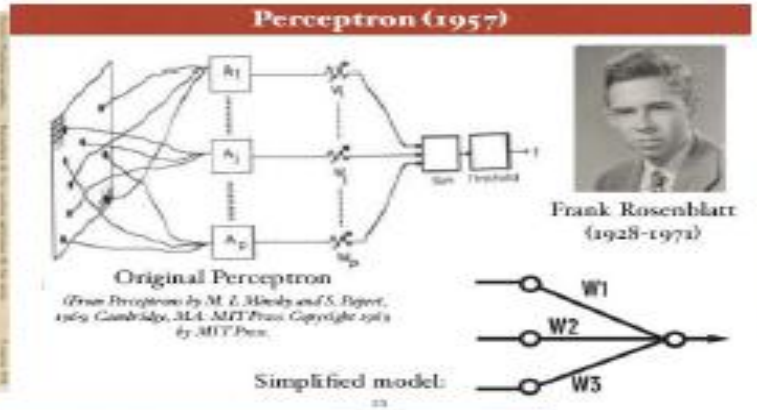


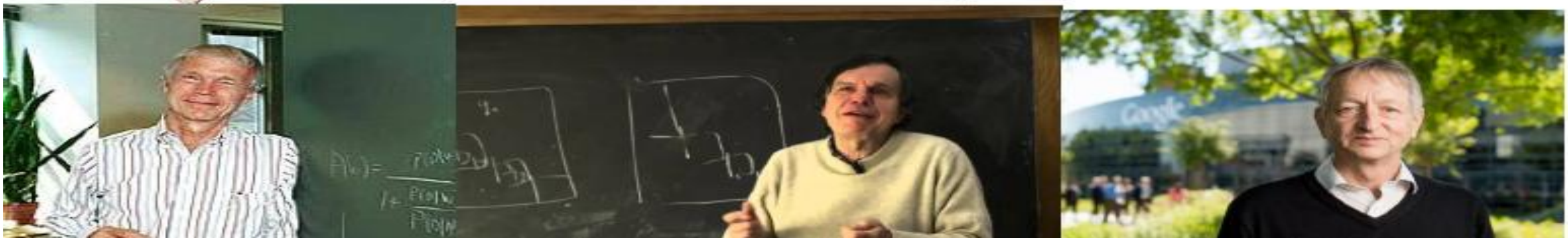
- James C. Maxwell (1831-1879) & D. Hilbert (1862-1943)-
- → B. Russell (1872-1970) A. Whitehead (1861-1947)
- K. Godel (1906-1978) → A. Turing, J. Von Neumann → G. Chaitin (1947-)
- Claude Shannon (1916-2001) Norber Weiner (1894-1964)





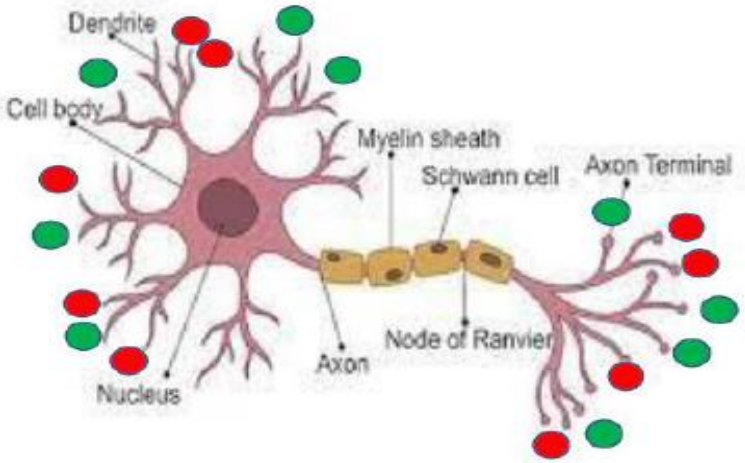






Before we start the course let's share a chat on what is its focus...

: UNA CHIACCHIERATA SULLE RETI NEURALI, CON UNA BREVE INTRODUZIONE STORICA, PER VEDERE COSA STIAMO FACENDO AL GIORNO D'OGGI.

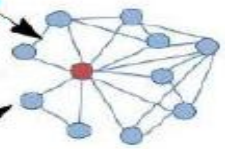


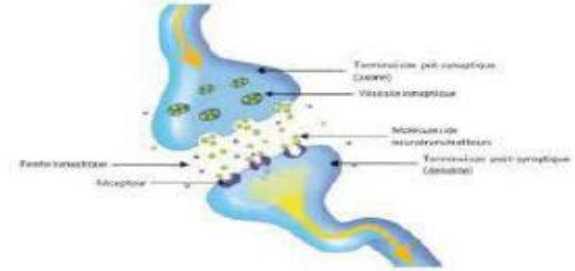
COGNITION

Learning
(dinamica sinaptica)
→ machine learning
→ (statistical inference)

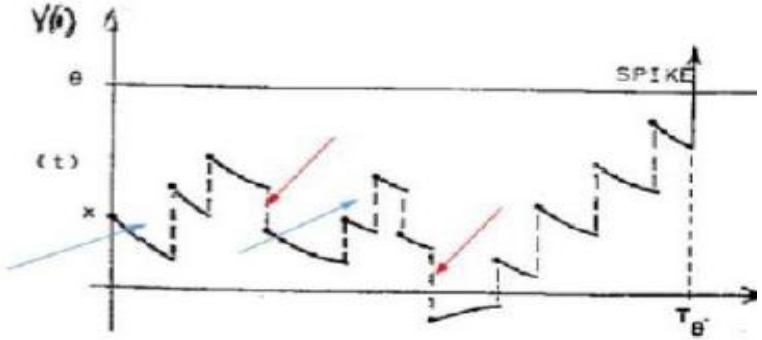
---Born Oppenheimer---

Retrieval
(dinamica neurale)
→ neural networks
→ (statistical physics)

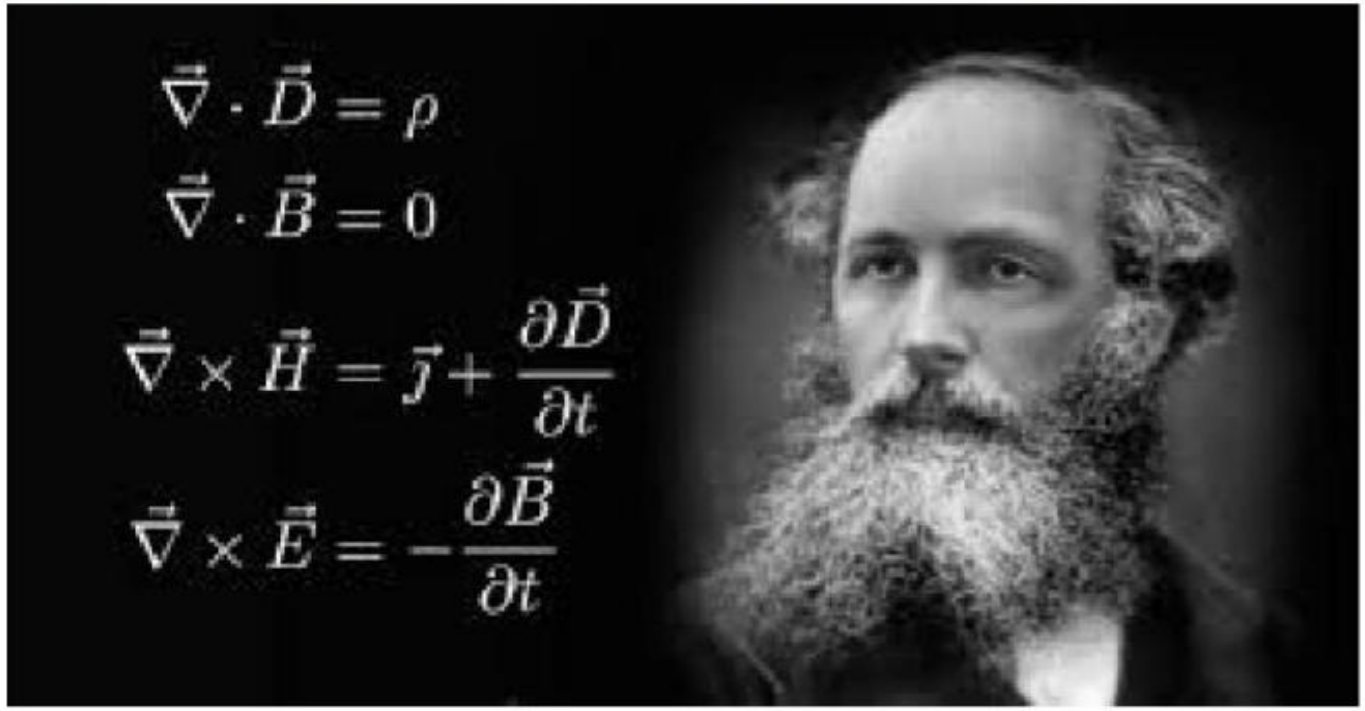


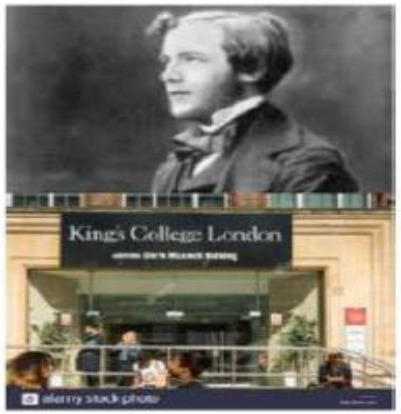


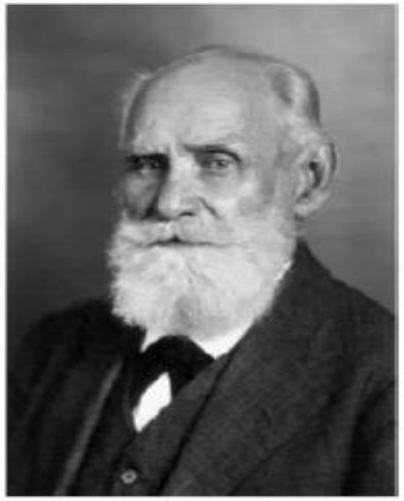
In realtà c'è molto di più... ma andiamo per ordini successivi...

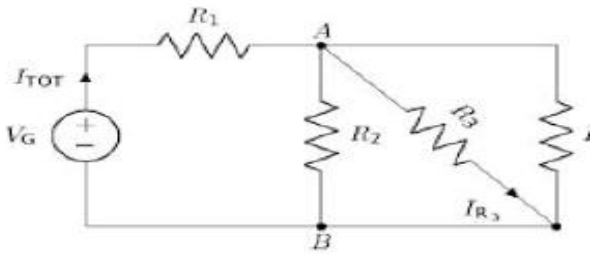


Un momento di sintesi: Maxwell & Pavlov
A dynamical theory of electromagnetic field
Anno di pubblicazione: 1865 (Pavlov Nobel 1904)
Il trionfo del Riduzionismo e della Scienza Dura







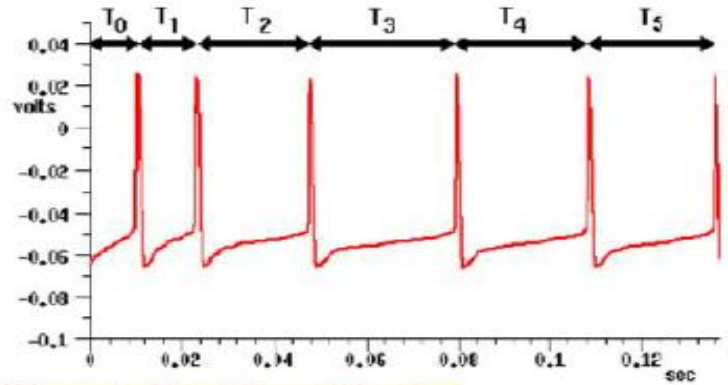


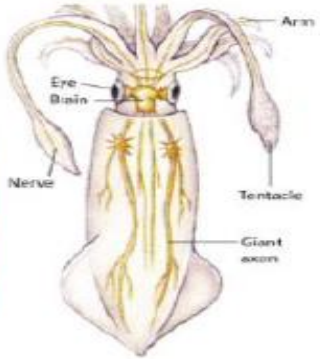
Nascono le reti elettriche (non elettroniche! per quelle bisogna aspettare la genesi dell'OpAmp, cioè metà del secolo scorso

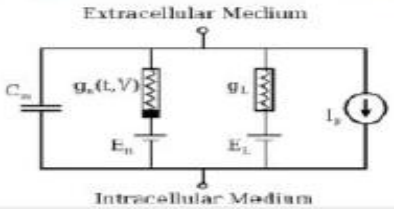
Nota: una conoscenza "moderna" della Fisica avviene antecedentemente a quella della Biochimica (per la quale servono strumenti fisici,e.g. microscopio ottico),

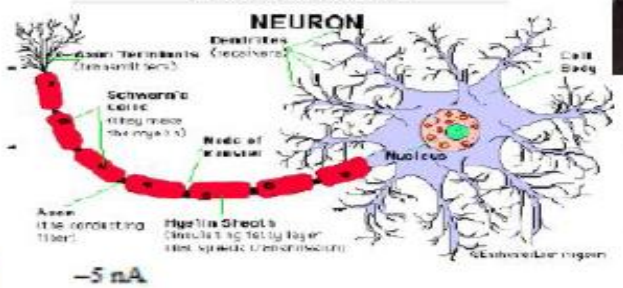
Le Reti Neurali (dei moduli corticali del nostro cervello) sono RETI ELETTRICHE... è un caso?

I neuroni nei calamari giganti attentano ai pilastri della Scienza: il Riduzionismo









Santiago Ramón y Cajal

From Wikipedia, the free encyclopedia

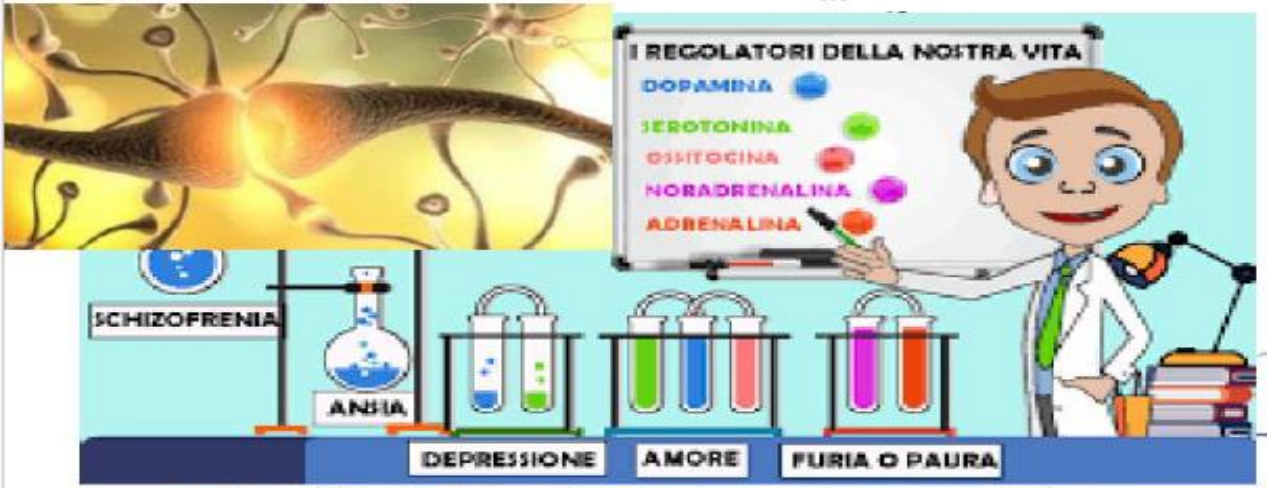
In this Spanish name, the first or paternal surname is Ramón and the second or maternal

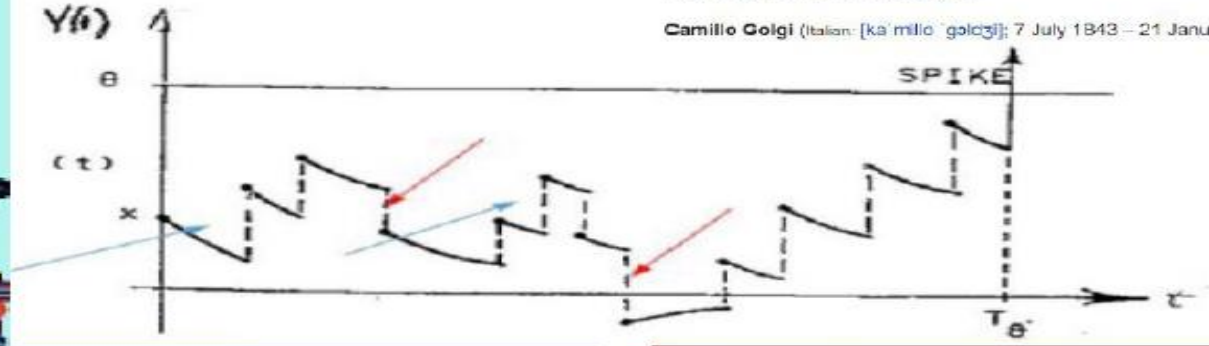
Santiago Ramón y Cajal (Spanish: [sanˈtjaɣo ramoˈn i kaˈxal]; 1 May 1852 – 17 October 193

Camillo Golgi

From Wikipedia, the free encyclopedia

Camillo Golgi (Italian: [kaˈmillo ˈɡɔlɡi]; 7 July 1843 – 21 January 192





Alan Hodgkin & Andrew Huxley ricevono il Nobel nel 1963

Golgi e Cajal ricevono il Premio Nobel nel 1906

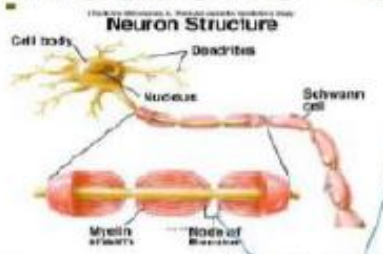
Capita la teoria delle reti elettriche (peraltro il secondo principio di Kirchhoff è sulle «Maglie» ma non è un concetto Bari-centrale) i matematici applicati si sono messi a studiare il soggetto elettrico biologico, cioè il neurone.

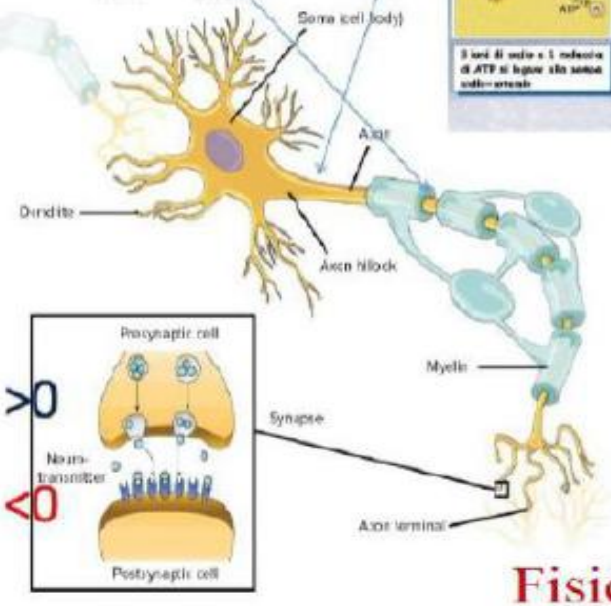
Notate che lo chiamo «soggetto elettrico biologico»: si sono studiati i soggetti del campo.

Primo ruolo cardine di Mat & FisTh: fornire strumenti quantitativi, metodi e modelli matematici per una descrizione "robusta" del fenomeno in esame (e risolverli anche)

Il Neurone «tipico»

Costa «lavoro» la pompa Na+/K+
...come negli Op-Amp (V_sat)



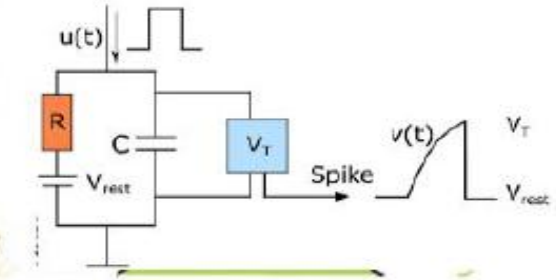


Nel mentre: uno di tre, anni 40→60

Il neurone "tipico" è "stupido": è un sommatore dispersivo a soglia.

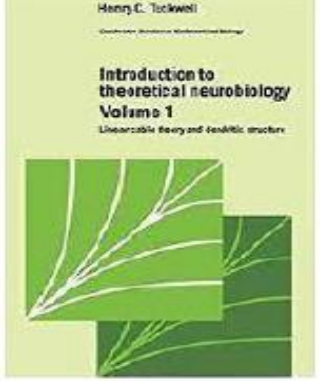
HH equations

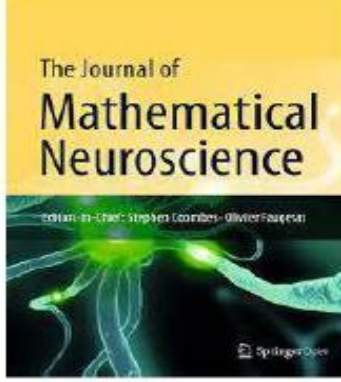
$$C\frac{dv}{dt} = I - g_{Na}m^3h(V-V_{Na}) - g_K n^4(V-V_K) - g_L(V-V_L)$$
$$\frac{dm}{dt} = a_m(V)(1-m) - b_m(V)m$$
$$\frac{dh}{dt} = a_h(V)(1-h) - b_h(V)h$$
$$\frac{dn}{dt} = a_n(V)(1-n) - b_n(V)n$$
$$a_m(V) = .1(V+40)/(1-\exp(-(V+40)/10))$$
$$b_m(V) = 4\exp(-(V+65)/18)$$
$$a_h(V) = .07\exp(-(V+65)/20)$$
$$b_h(V) = 1/(1+\exp(-(V+35)/10))$$
$$a_n(V) = .01(V+55)/(1-\exp(-(V+55)/10))$$
$$b_n(V) = .125\exp(-(V+65)/80)$$



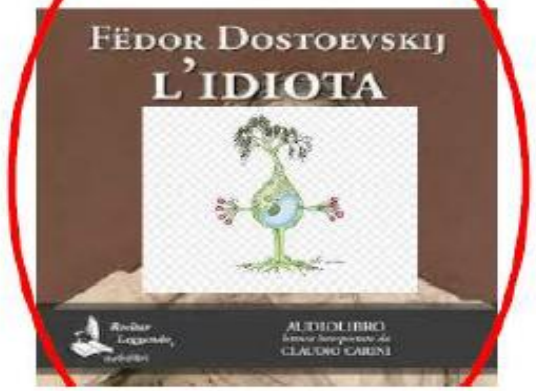
Fisica Statistica!

ALLA RICERCA DI "PROPRIETA' EMERGENTI"





Jülich Supercomputing Centre (JSC)
Website Directions Save
4.6 ★★★★★ 15 Google reviews
Research institute in Jülich, Germany

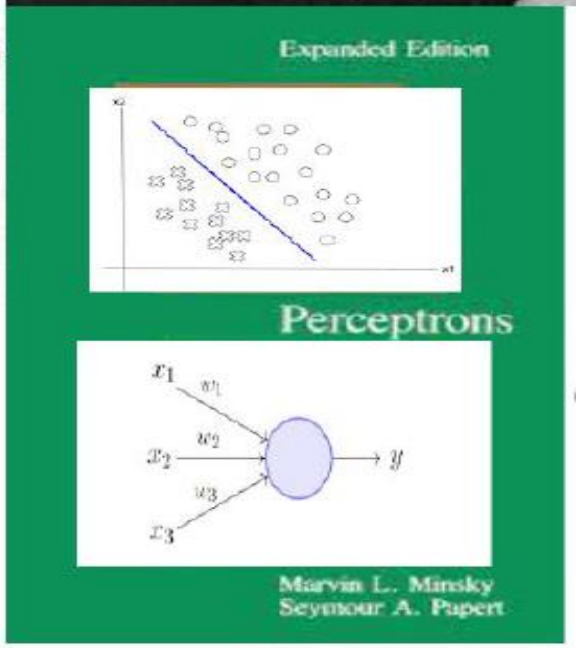


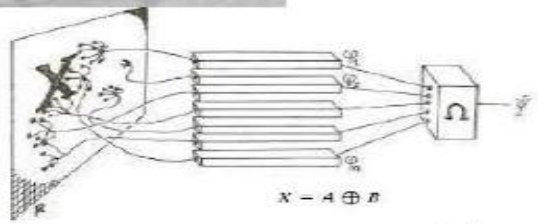
Qui entra in gioco la Meccanica Statistica ed in particolare la Teoria di Parisi (Vide Infra)

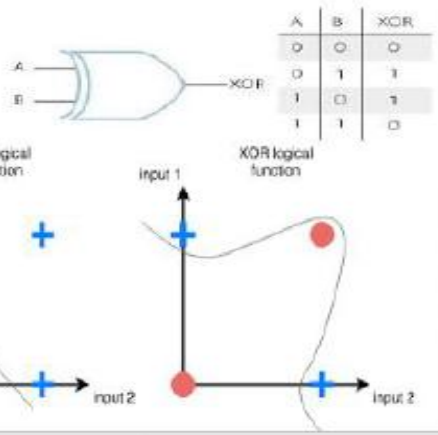
Anni 60/70: il winter time dell'AI, la nascita della Computer Science e della Macchina di Turing: PC ed AI si separeranno per 40 anni, nei quali ci sarà una «rivoluzione silenziosa»

Intanto i futuri «Computer Scientists» (ignari della MS)...

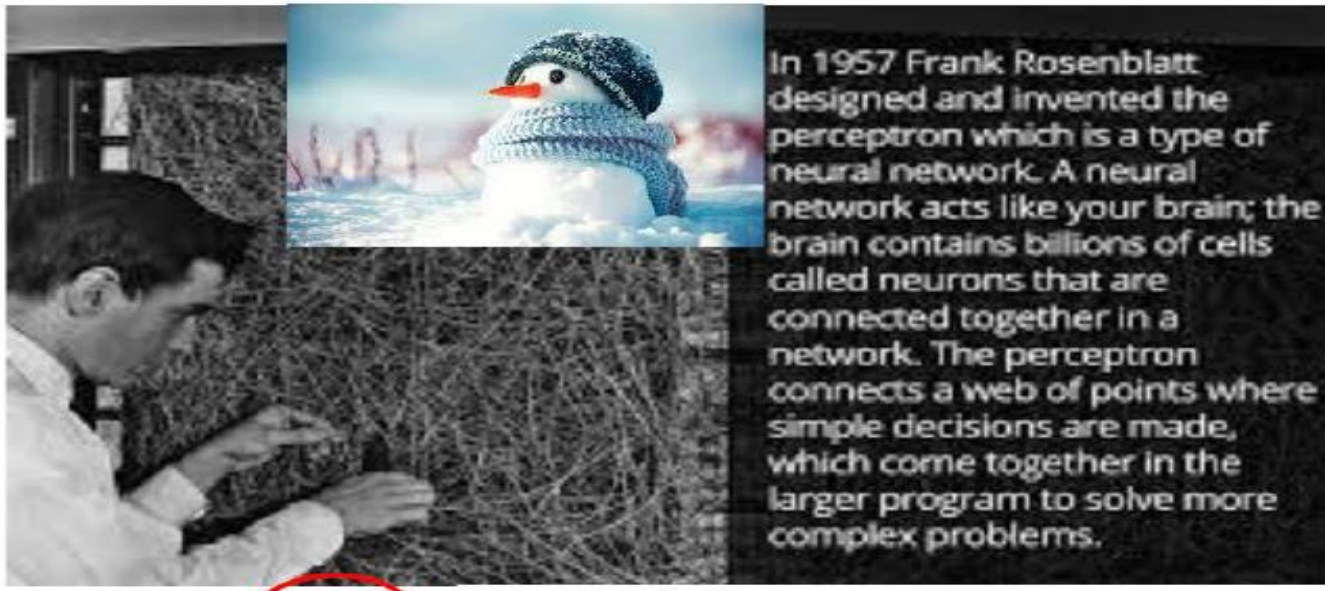
Il winter time: gli anni 70



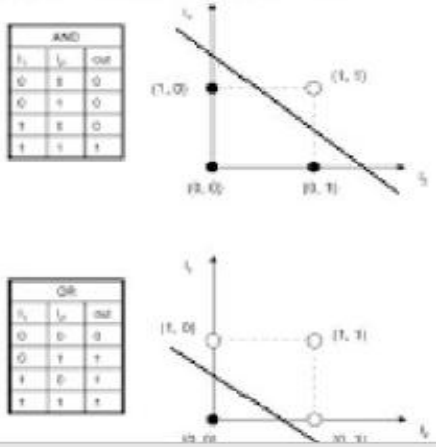




A	B	XOR
0	0	0
0	1	1
1	0	1
1	1	0



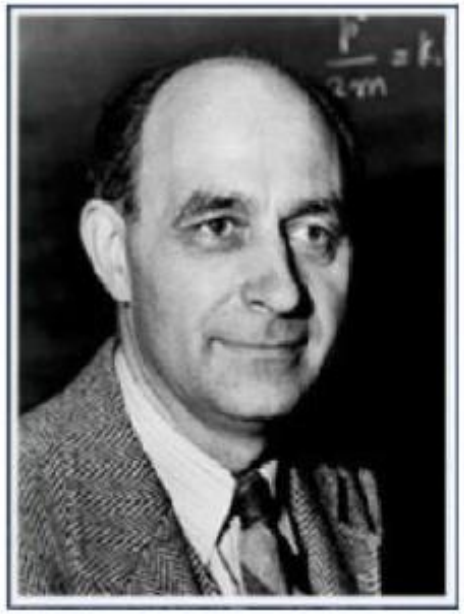
1958 New York Times...



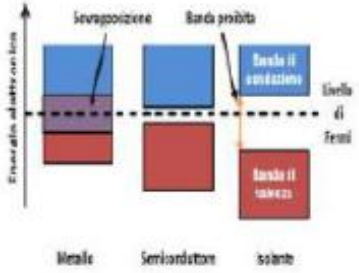
La genesi dell'Op-Amp da una prima accelerata (via le rotative): nascono le macchine deterministiche (le macchine di Turing), i PC ma fallisce il primo tentativo di un'AI. (un'altra accelerata avviene nel 2000 con lo shift CPU->GPU)

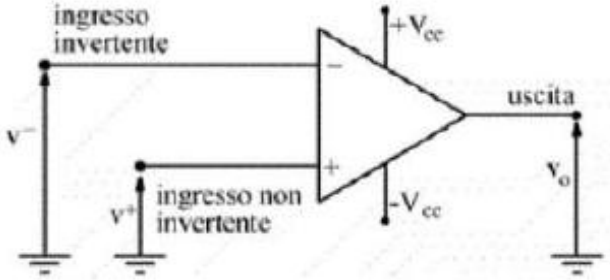
Nel mentre: due di tre: anni 30 → 50

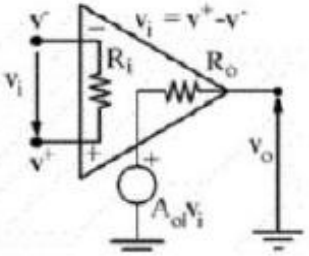
Nel mentre (tra il 1900 ed il 1950) nasce la Meccanica Quantistica:
→ decolla lo studio di dei semiconduttori, le rotative abdicano ai chip

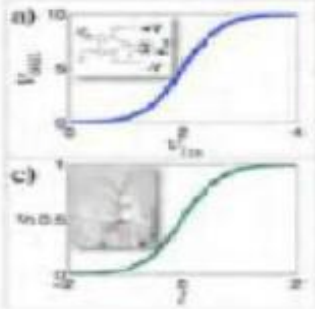


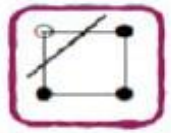
Il ruolo dei Matematici è stato enorme anche nella genesi della MQ





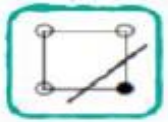




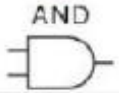


OR
(v)

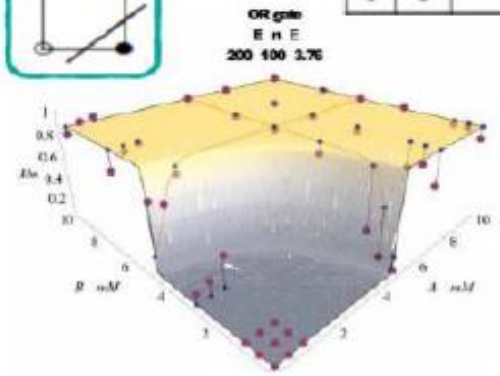
AND
(^)

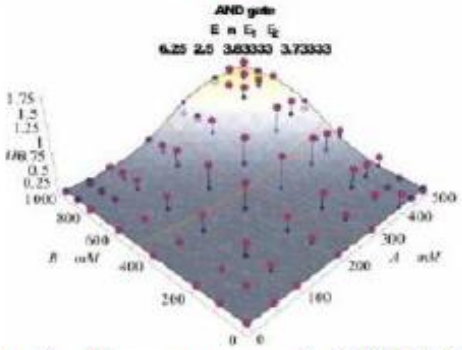


AND



INPUT		OUTPUT
A	B	
0	0	0
1	0	0
0	1	0
1	1	1





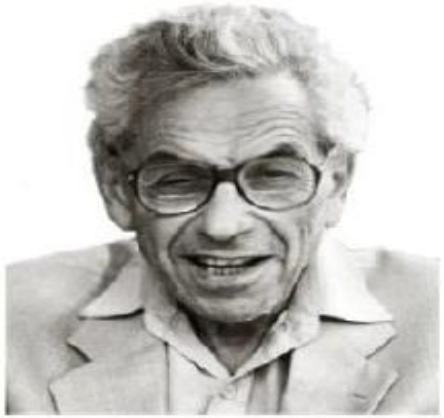
clock di neurone è O(10^3)
clock di CPU è (10^9)
il rapporto è un milione...

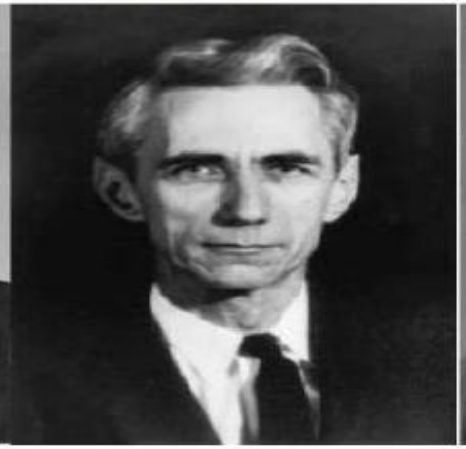
Di cosa si tratta? Di un fatto ovvio ma per nulla banale… e sta ri-scrivendo in todo la Scienza (è uno dei tre concetti non ovvii che vorrei veicolarvi, un altro è sul «Riduzionismo» e l'ultimo sulla «necessità dell'elettricità?» ma li vediamo alla fine)

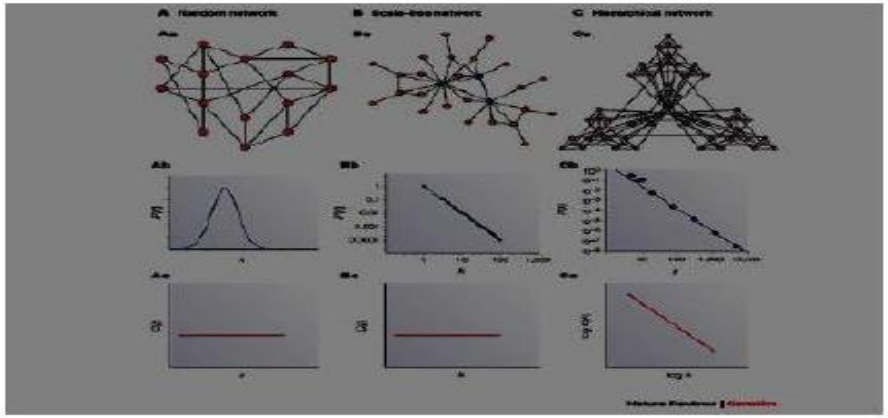
Una rivoluzione silenziosa: L'attenzione si sposta dai soggetti alle loro interazioni:
→ decolla lo studio di Teoria dei Grafi e di Teoria dell'Informazione (tutto Booleano)

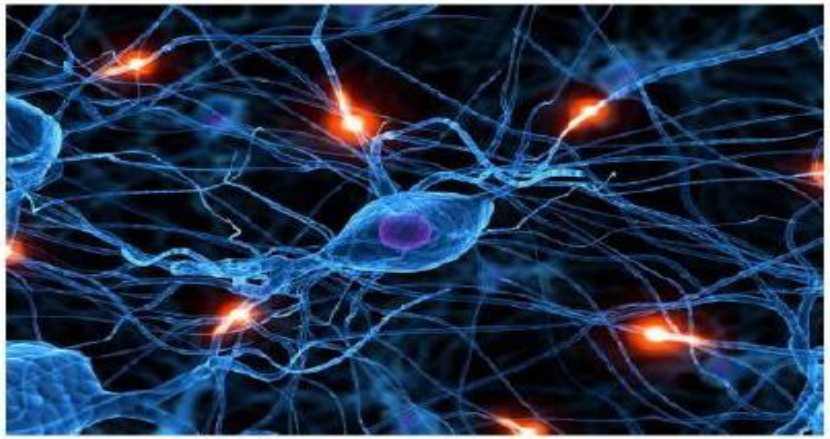
Ps) un neurone non sa fare lo XOR ma un pugnetto di neuroni si!

Nel mentre: due di tre: anni 60 → oggi

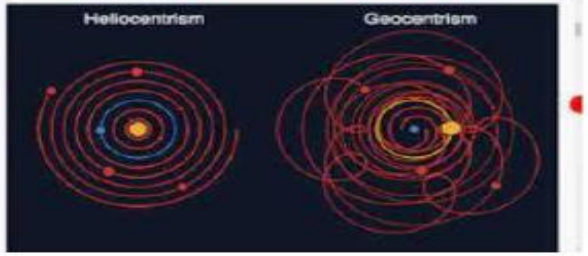




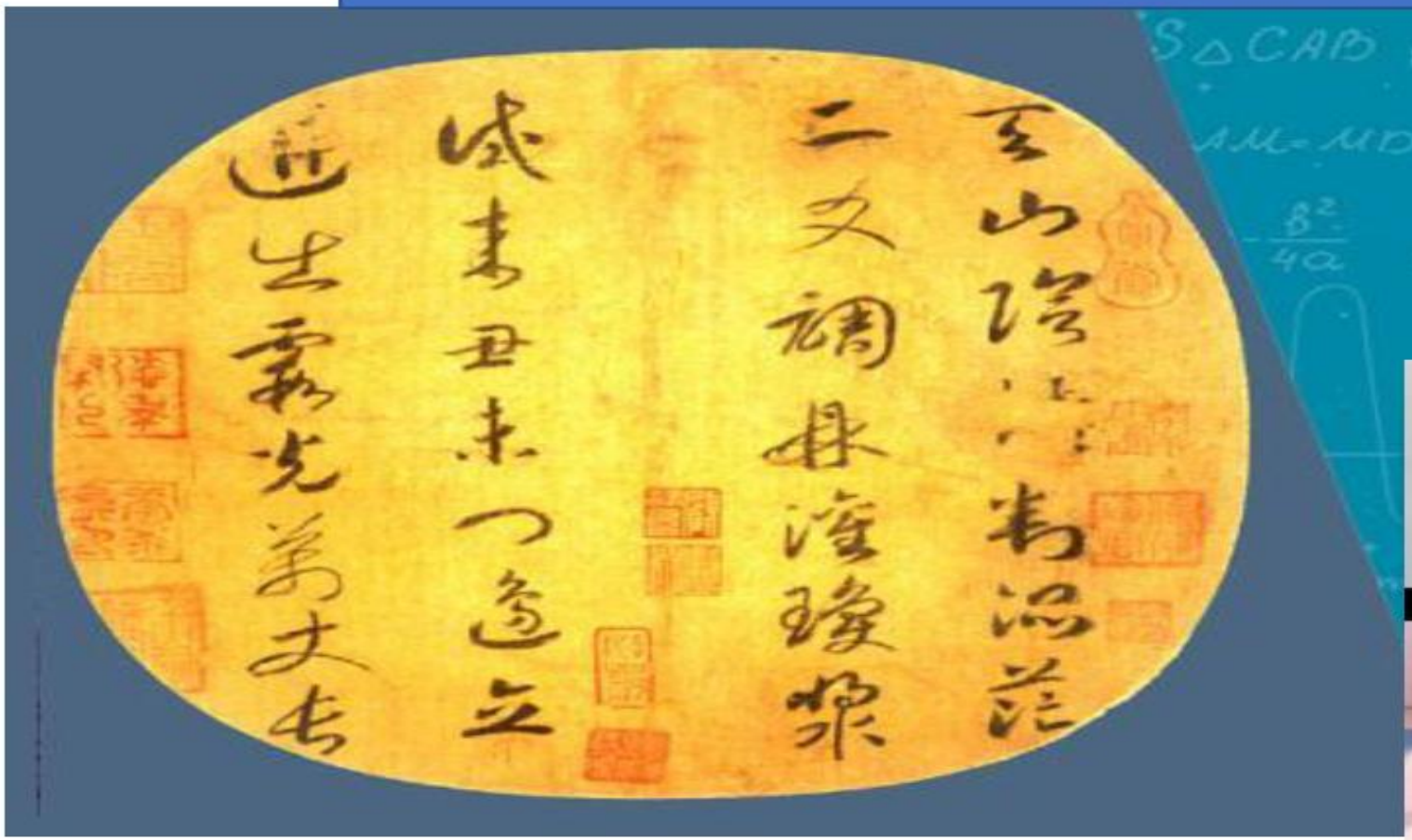




Per me questa è una silenziosa rivoluzione strutturale molto profonda



ANNI 80/90



J. Phys. A: Math. Gen. 19 (1986) L675–L680. Printed in Great Britain

LETTER TO THE EDITOR

Asymmetric neural networks and the process of learning

Giorgio Parisi
Dipartimento di Fisica, II Universita' di Roma 'Tor Vergata', Via Orazio Raimondo, Roma 00173, Italy and INFN, Sezione di Roma, Italy

J. Phys. A: Math. Gen. 19 (1986) L617–L620. Printed in Great Britain

LETTER TO THE EDITOR

A memory which forgets

Giorgio Parisi
Dipartimento di Fisica, II Università di Roma 'Tor Vergata', Via Orazio Raimondo, Roma 00173, Italy and INFN, sezione di Roma, Italy

IA tradizionale: un insieme di regole

Se l'oggetto ha
- **4 gambe**

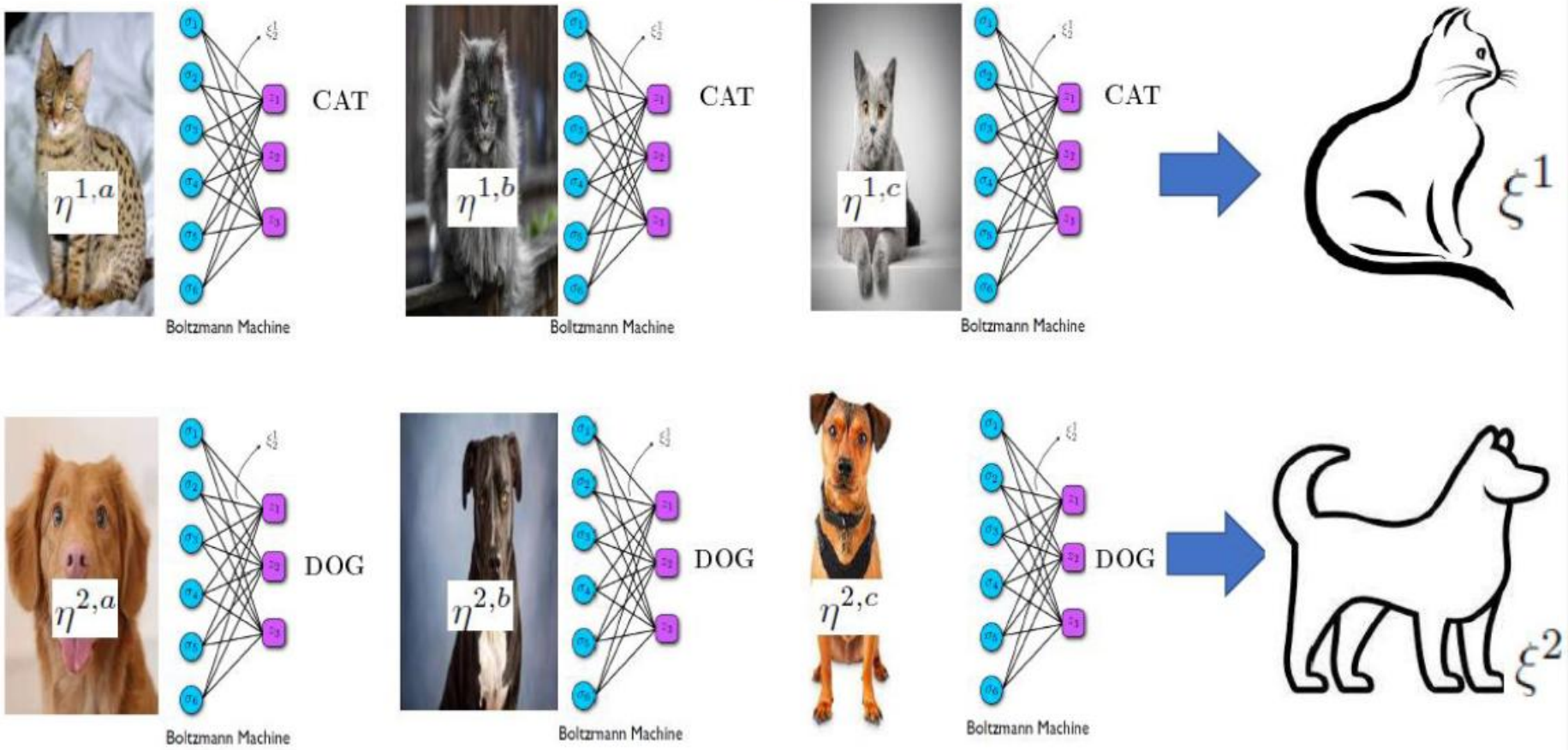
OR 5 rotelle

OR gambe ricurve ...

OR ...

L'archetipo di «gattino» o «cagnolino» può anche non esistere nella realtà (è un concetto dell'Iperuranio Platonico più che del contingente, come l'oscillatore armonico per la Fisica)

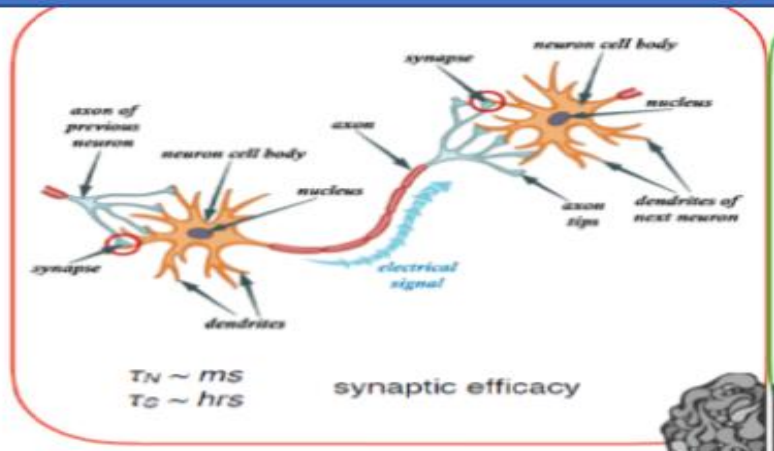
$$\Delta\xi_i^\mu = \epsilon\beta\left(\langle\sigma_i z_\mu\rangle_+ - \langle\sigma_i z_\mu\rangle_-\right)$$

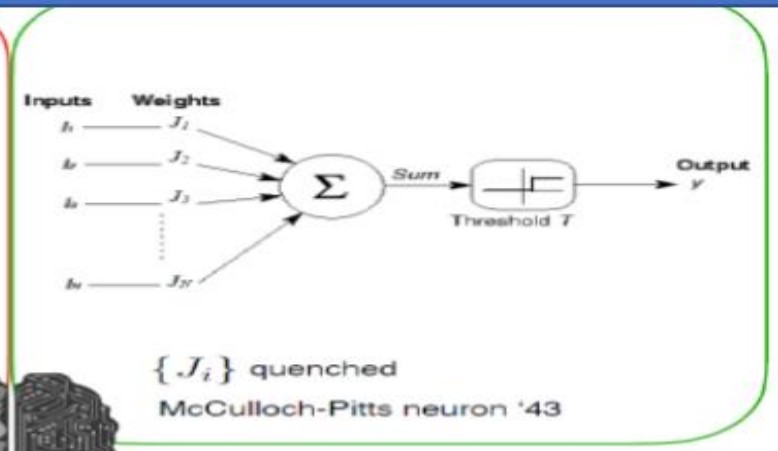


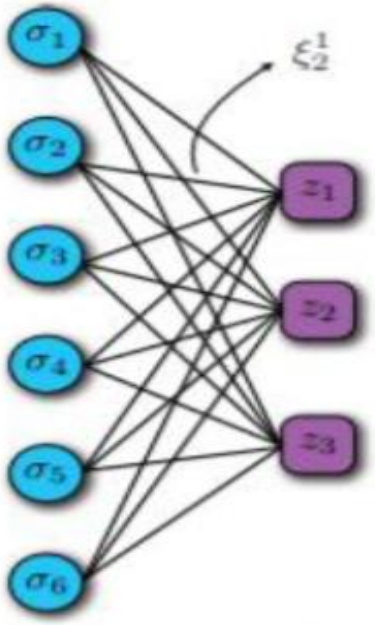
From the '80 the Second AI program starts as in the '70 the First AI program dies.

In the Grand-Mother cell setting we use one hidden neuron to label each archetype/pattern we want the network to learn/store/retrieve

Come avviene (negli anni 70/80) il cambio di paradigma, dal neurone alle reti di neuroni? Sempre ancorati all'ispirazione biologica, ora a livello di reti neurali (shallow)

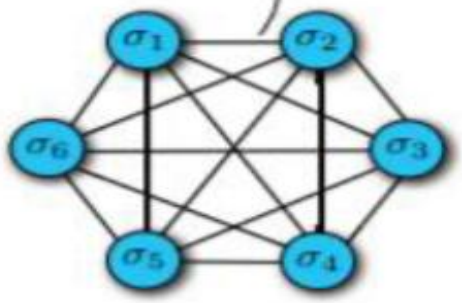






Boltzmann Machine

$$J_{12} = \sum_{\mu=1}^{3} \xi_1^\mu \xi_2^\mu$$



Hopfield Network

Come nella prima ondata abbiamo
dato luogo, come archetipo, al «Perceptrone»
nella seconda l'archetipo è la «RBM»:
Abbiamo sistematicamente spostato
L'attenzione dai soggetti (neuroni e sinapsi)
Alle reti che questi compongono e le proprietà
Emergenti delle stesse sono il focus dello
Springer Time dell'AI che stiamo vivendo.

RBM incastonate fanno DBM...

Results in DNN

The Computer Scientist approach is totally empirical: try & check… we need to buld a Theory for AI

(2010) speech recognition

(2012) object recognition in complex images

(2013) molecular biology for personalised medicine

(2013) industries set in

(2014) automatic translation, text to text, image to text

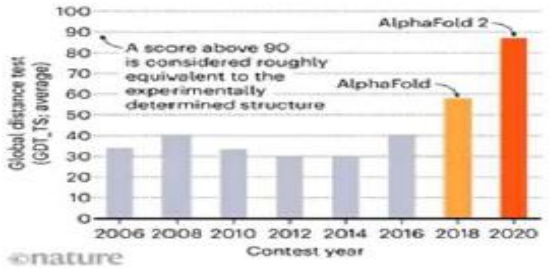
(2015) super-resolution, generators of realistic images, music, paintings …

(2016) play games using with strategic intuition

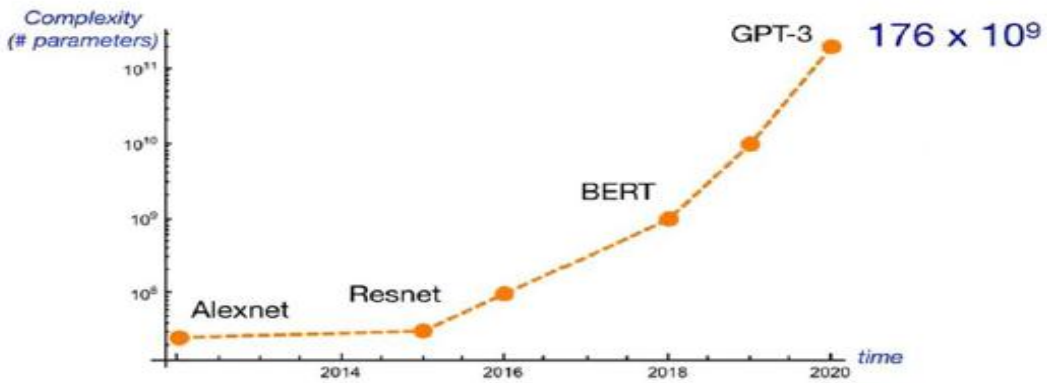
- > 2012 … applications across sciences: data from accelerators, new molecules, new drugs, astrophysics, solvers for PDEs, …

Latest: 2020

- GPT-3 (language processing)
- Alphafold (protein structure prediction from sequence data)



The growing complexity of DNN



Come funzione l'apprendimento supervisionato?

Si usano esempi annotati da umani (apprendimento supervisionato)

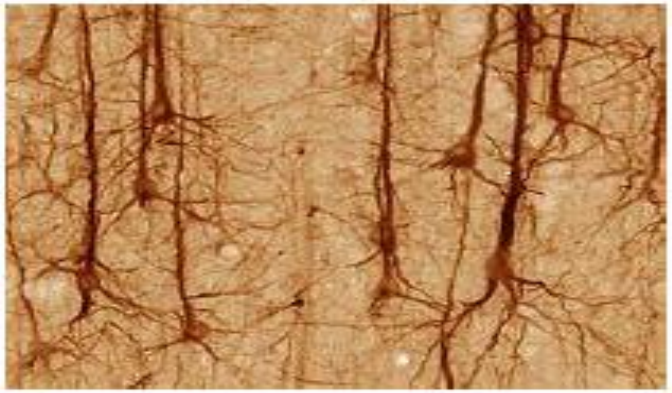
1. *Mostriamo degli esempi (input+output)*
2. *Se la rete sbaglia l'output, le connessioni **(milioni)** vengono modificate automaticamente*
3. *Ripetiamo **milioni** di volte su **milioni** di esempi (training set)*
4. *Alla fine, quando la rete ha imparato a classificare correttamente gli esempi, possiamo iniziare ad usarla per fare predizioni. Viene*

Una prima introduzione/chiacchierata/quadro generale

WHY ARTIFICIAL NEURAL NETWORKS?

The long course of evolution has given the human brain many desirable characteristics not present in von Neumann or modern parallel computers. These include

- massive parallelism,
- distributed representation and computation,
- learning ability,
- generalization ability,
- adaptivity,
- inherent contextual information processing,
- fault tolerance, and
- low energy consumption.



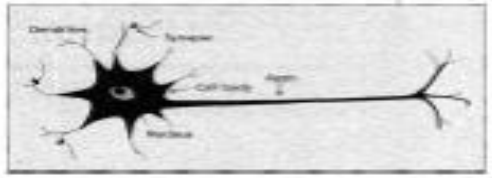
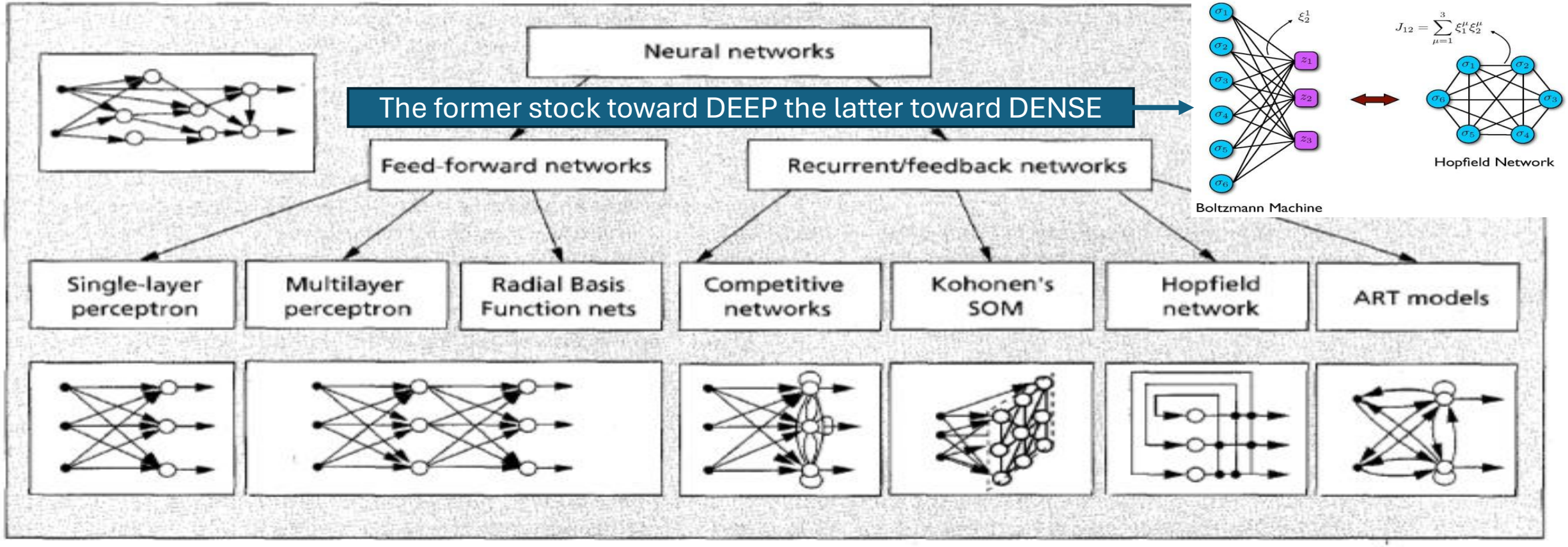


Table 1. Von Neumann computer versus biological neural system.

	Von Neumann computer	Biological neural system
Processor	Complex High speed One or a few	Simple Low speed A large number
Memory	Separate from a processor Localized Noncontent addressable	Integrated into processor Distributed Content addressable
Computing	Centralized Sequential Stored programs	Distributed Parallel Self-learning
Reliability	Very vulnerable	Robust
Expertise	Numerical and symbolic manipulations	Perceptual problems
Operating environment	Well-defined, well-constrained	Poorly defined, unconstrained

Una prima introduzione/chiacchierata/quadro generale



Learn

$3 R1&5C!

4 L3GG&R3

QÜ3S70 N()N

$E1 NOЯM4L3

Retrieve

Struttura del corso

Intro: non tecnico (merging di pre-requisiti specifici)

Main body A: Machine Retrieval. Tecniche di Meccanica Statistica [direct problem, neural dynamics]

Modello 1) CW: archetipo di «sistema semplice» 3 tecniche (Guerra, PDE, ReplicaTrick)
Modello 2) Hopfield low storage [low storage statistical mechanics = low D statistical inference] (Guerra, PDE, ReplicaTrick)
Modello 3) SK: archetipo di «sistema complesso» 3 tecniche (Guerra, PDE, Replica Trick)
Modello 4) Hopfield high storage [high storage statistical mechanics = high D Stat. Inf.] (Guerra, PDE, ReplicaTrick)

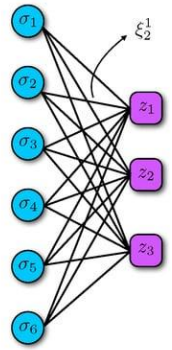
Main body B: Machine Learning. Tecniche di Inferenza Statistica [inverse problem,synaptic dynamics]

Modello 1) Perceptrone, minimi quadrati, etc.
Modello 2) Hinton's RBM con Teacher: Supervised Learning
Modello 3) Hinton's RBM con Teacher: Unsupervised Learning
Modello 4) Hebbian learning and duality artificial & biological.

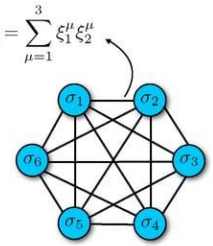
Outro: ora che abbiamo l'»oscillatore armonico» per orientarci, spaziamo in tante variazioni sul tema.
e.g. Problema del commesso viaggiatore, Reti Multitasking, Reti Dreaming, Reti dense e profonde, ,etc.

Learn

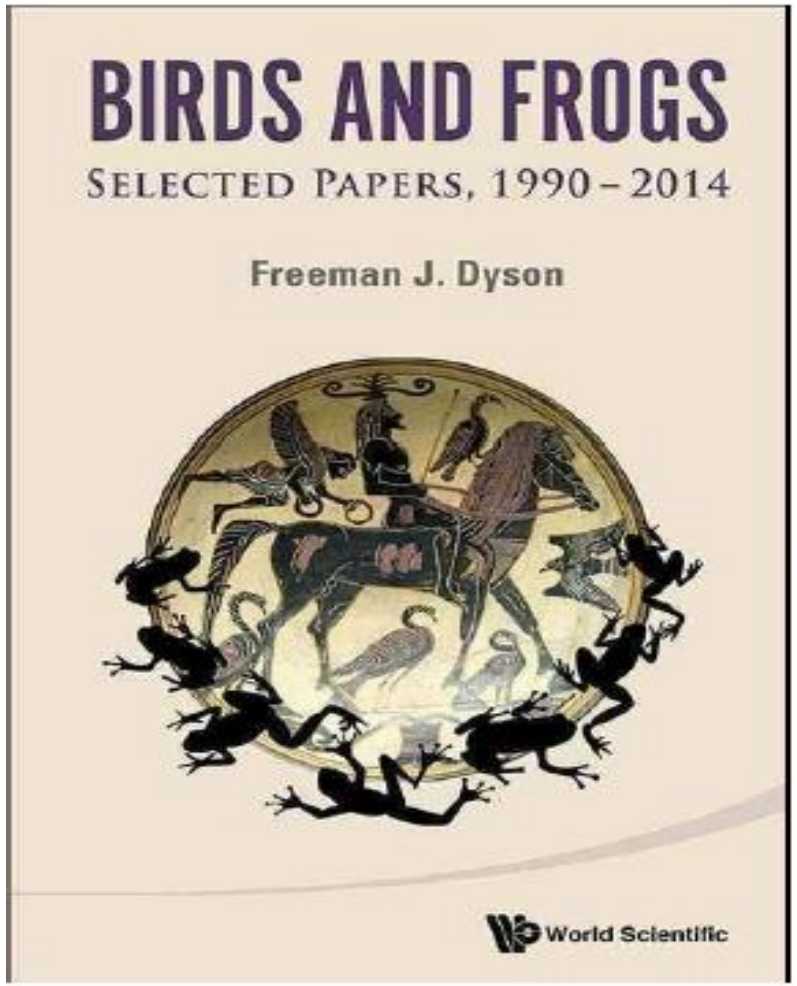
Retrieve

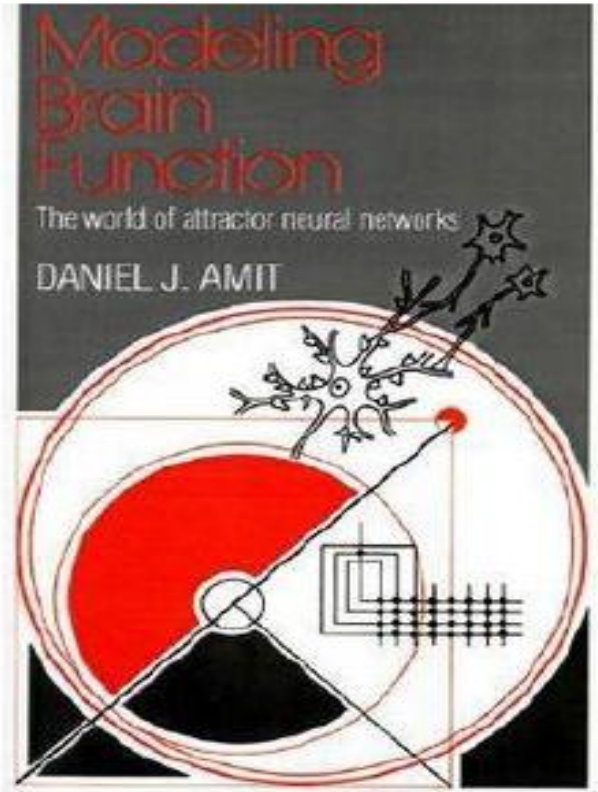


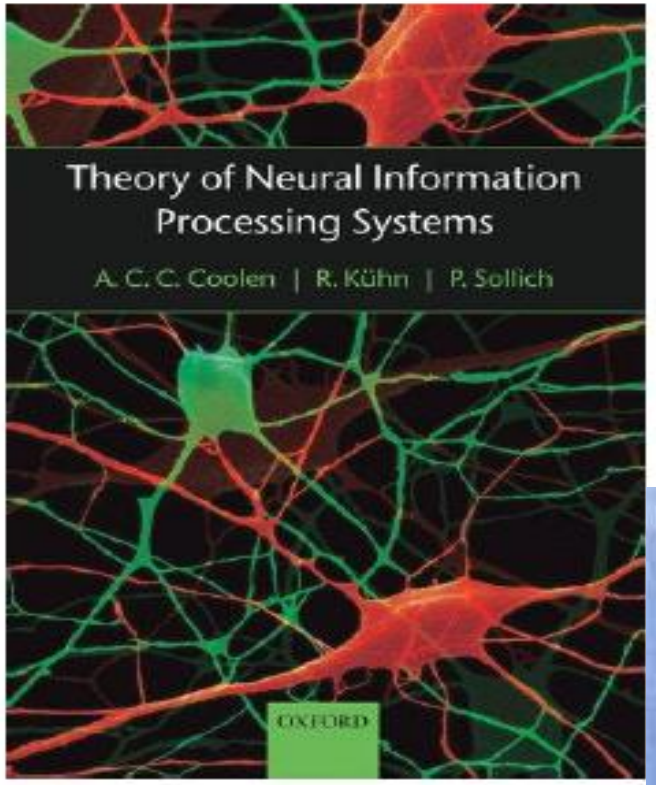
Boltzmann Machine

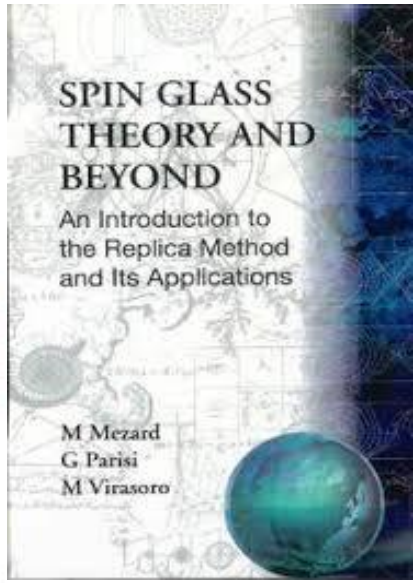


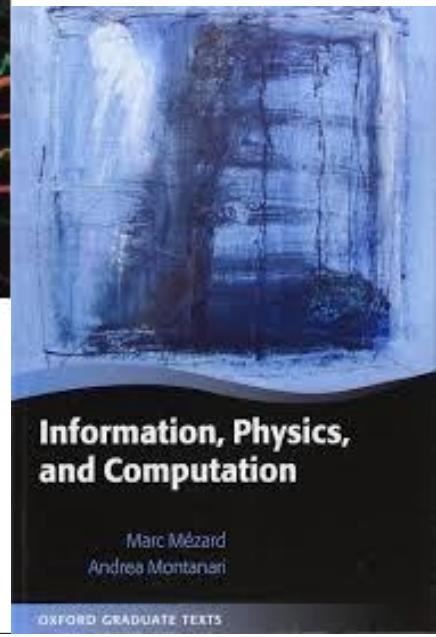
Hopfield Network


BIRDS AND FROGS
SELECTED PAPERS, 1990 – 2014
Freeman J. Dyson
World Scientific


Modeling Brain Function
The world of attractor neural networks
DANIEL J. AMIT


Theory of Neural Information Processing Systems
A. C. C. Coolen | R. Kühn | P. Sollich
OXFORD


SPIN GLASS THEORY AND BEYOND
An Introduction to the Replica Method and Its Applications
M Mezard
G Parisi
M Virasoro


Information, Physics, and Computation
Marc Mézard
Andrea Montanari
OXFORD GRADUATE TEXTS

MIT Technology Review

MIT Technology Review

Artificial intelligence / Machine learning

Training a single AI model can emit as much carbon as five cars in their lifetimes

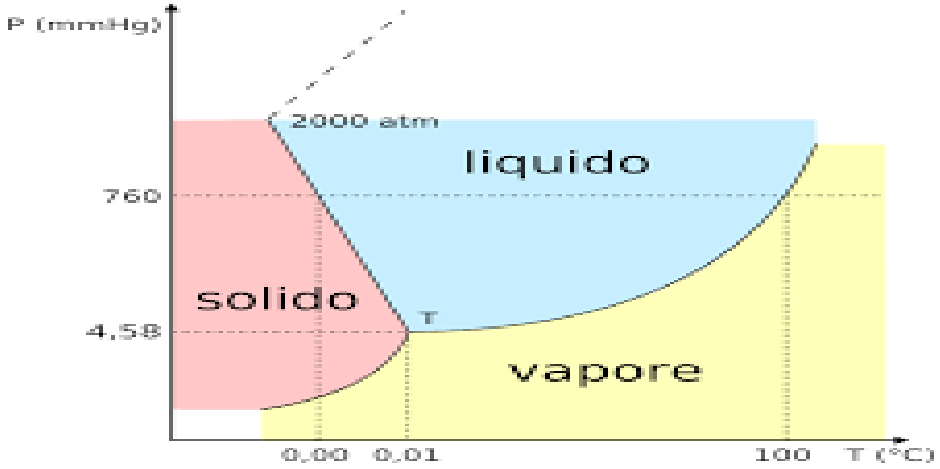
Deep learning has a terrible carbon footprint.

by **Karen Hao** June 6, 2019

A data center

MECCANICA QUANTISTICA MODERNA

MECCANICA STATISTICA MODERNA



L'assunto di «un mondo a campana» è un antico ed annoso retaggio della prospettiva riduzionista che nelle scienze non-dure si è rivelato un fallimento colossale!

PNAS

Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates

Anders Eklund[a,b,c,1], Thomas E. Nichols[d,e], and Hans Knutsson[a,c]

[a]Division of Medical Informatics, Department of Biomedical Engineering, Linköping University, S-581 85 L
Machine Learning, Department of Computer and Information Science, Linköping University, S-581 83 Lin
Science and Visualization, Linköping University, S-581 83 Linköping, Sweden; [d]Department of Statistics, U
Kingdom; and [e]WMG, University of Warwick, Coventry CV4 7AL, United Kingdom

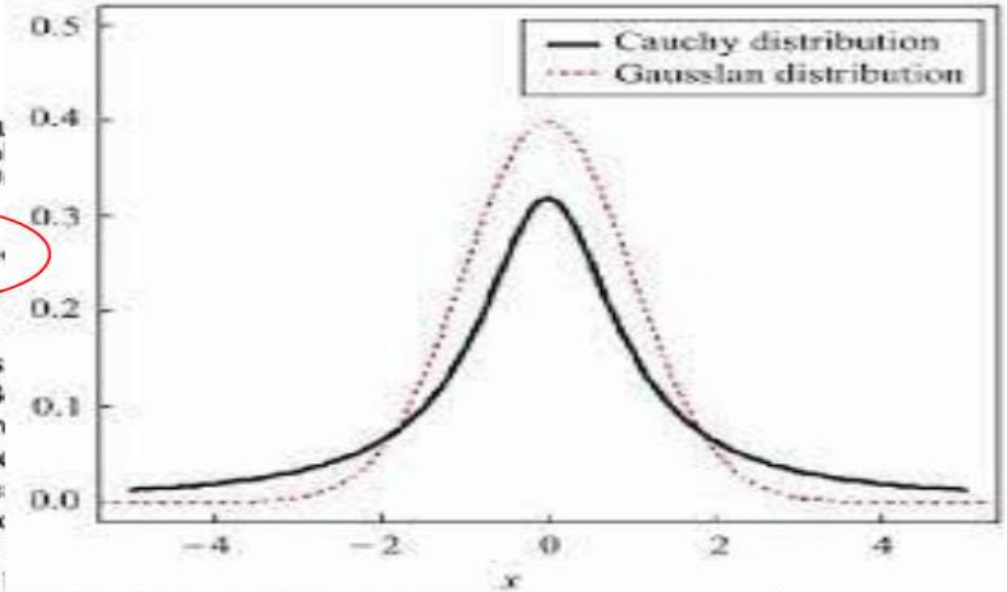
Edited by Emery N. Brown, Massachusetts General Hospital, Boston, MA, and approved May 17, 2016 (re

The most widely used task functional magnetic resonance imaging (fMRI) analyses use parametric statistical methods that depend on a variety of assumptions. In this work, we use real resting-state data and a total of 3 million random task group analyses to compute empirical familywise error rates for the fMRI software packages SPM, FSL, and AFNI, as well as a nonparametric permutation method. For a nominal familywise error rate of 5%, the parametric statistical methods are shown to be conservative for voxelwise inference and invalid for clusterwise inference. Our results suggest that the principal cause of the invalid cluster inferences is spatial autocorrelation functions that do not follow the assumed Gaussian shape. By comparison, the nonparametric permutation test is found to produce nominal results for voxelwise as well as clusterwise inference. These findings speak to the need of validating the statistical methods being used in the field of neuroimaging.

fMRI | statistics | false positives | cluster inference | permutation test

(FWE), the chance of
measure the FWE as
any significant results
one-sample designs. B
drawn from a large gr
of no group differenc
over, because the re
consistent shifts in blo
for a single group of
activation should also
voxelwise inference, where significance is individually assessed at each voxel, and clusterwise inference (19–21), where significance is assessed on clusters formed with an arbitrary threshold.

In brief, we find that all three packages have conservative voxelwise inference and invalid clusterwise inference, for both one- and two-sample t tests. Alarmingly, the parametric methods



Come faccio «un tanto al chilo» a mettermi in guardia da queste distribuzioni?

Da dove ha origine questo retaggio Gaussiano? Può non sembra ma… dal Riduzionismo! →

le code larghe (non si vedono) ma l'integrale deve fare 1 e la media è molto piccata (c'è poca area lì sotto)…

Il paracadute di Aristotele come archetipo di struttura di rivoluzione scientifica

(discuti anche la relatività ristretta che si capisce)

Metafora che, declinata nel mondo dell'inferenza ad alta dimensionalità, significa

«se il sistema è semplice la statistica collassa su quella classica (Gaussiana)»

Osservazione utile per AI dopo: che stiamo facendo?

Impariamo un concetto in un contesto e poi lo esportiamo in altri contesti (memoria associativa CAM)

$$F = ma. \quad F_1 = mg, \quad F_2 = -\beta v = -\beta \frac{dx}{dt}, \quad ma = m\frac{dv}{dt}, \tag{0.1}$$

$$mg - \beta v = ma \rightarrow mg - \beta \frac{dx}{dt} = m\frac{dv}{dt} \tag{0.2}$$

$$g - \frac{\beta}{m}v = \frac{dv}{dt} \rightarrow g - \frac{\beta}{m}v = \frac{dv}{dt} \tag{0.3}$$

$$dt = \frac{dv}{g - \frac{\beta}{m}v} \rightarrow \int dt = \int \frac{dv}{g - \frac{\beta}{m}v} \tag{0.4}$$

$$\tag{0.5}$$

$$\int_{t_0}^{t} dt = \int_{v_0}^{v(t)} \frac{dv}{g - \frac{\beta}{m}v}$$

pongo $y = g - \frac{\beta}{m}v$ tale che $dy = -\frac{\beta}{m}dv$ e e quindi $dv = -\frac{m}{\beta}dy$

$$(t - t_0) = -\frac{m}{\beta}\int_{y_0}^{y(t)} \frac{dy}{y} \rightarrow -\frac{\beta}{m}t = \ln(y(t)) - \ln(y_0) = \ln(\frac{y(t)}{y_0}). \tag{0.6}$$

quindi $y(t) = y_0 e^{-\frac{\beta}{m}t}$ e $lim_{t\rightarrow\infty} y(t) = 0$ quindi $y = g - \frac{\beta}{m}v = 0$ e quindi

$$v_{limite} = \frac{mg}{\beta} \tag{0.7}$$

Discutilo anche dal punto di vista TD

v_{limite} e' costante (ed ha anche senso, giocate con le variabili).

Come diceva Aristotele a forza costante corrisponde velocita' costante fuori dal mondo di Platone. E, noi capiremo alla fine della settimana che la rivoluzione moderna del Deep Learning collassa sul Teorema del Limite Centrale per le reti shallow nei mondi semplici.